Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Video-Based Unsupervised Ergonomic Posture Assessment in Motor-Manual Forest Operations Using Deep Autoencoders and the OWAS Framework

Gabriel Osei Forkuo

gabriel.forkuo@unitbv.ro

Transilvania University of Brasov

Frank Adomako-Kwabia

The Forestry Commission of Ghana

Research Article

Keywords: Forestry ergonomics, OWAS, motor-manual operations, pose estimation, unsupervised learning, autoencoder, anomaly detection, occupational safety

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11823232/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Video-Based Unsupervised Ergonomic Posture Assessment in Motor-Manual Forest Operations Using Deep Autoencoders and the OWAS Framework

Gabriel Osei Forkuo[1, *] and Frank Adomako-Kwabia[2]

[1]Department of Forest Engineering, Forest Management Planning and Terrestrial Measurements, Faculty of Silviculture and Forest Engineering, Transilvania University of Brasov, Şirul Beethoven 1, 500123, Brasov, Romania, gabriel.forkuo@unitbv.ro; https://orcid.org/0000-0001-8478-8066 (G.O.F.)

[2]Forestry Commission of Ghana, Forest Services Division P. O. Box GP 257, Accra-Ghana, fakwabia.fsd@fcghana.org; https://orcid.org/0009-0001-1911-6236 (F.A-K)

[*]**Corresponding author:** gabriel.forkuo@unitbv.ro (G.O.F.)

Abstract

Background: Motor-manual forest operations, in which workers fell, delimb, and cross-cut standing timber with handheld chainsaws, remain physically hazardous, yet conventional observational ergonomic methods such as the Ovako Working Posture Analysing System (OWAS) cannot feasibly cover full shifts across many workers because they rely on labour-intensive manual coding. This study aimed to design and validate a fully unsupervised, video-based pipeline that quantifies OWAS ergonomic risk exposure in real motor-manual forestry footage without any manually labelled ground truth.

Methods: We analysed 36,783 pose-estimated frames from 35 field recordings of chainsaw operators using MediaPipe PoseLandmarker, converted skeletal coordinates into seven biomechanical features, and partitioned the data by video identity (GroupShuffleSplit) to prevent temporal leakage. Classical detectors (Isolation Forest, One-Class Support Vector Machine, and PCA with K-Means) and a compact symmetric deep autoencoder learned normative posture structure. Three ablation studies, clustering-validity indices,

and Cohen's kappa were used to interrogate robustness before cluster centroids were mapped to four-digit OWAS codes and Action Categories (AC).

Results: The autoencoder converged stably with a four-dimensional bottleneck (validation mean squared error = 0.089 in the original run; 0.082 ± 0.002 across five seeds of an independent replication). Agreement between the Isolation Forest and the autoencoder on the held-out split was moderate (Cohen's kappa = 0.431); applying both criteria to the full corpus identified 777 frames (2.11%) as consensus anomalies. Three posture archetypes emerged (silhouette = 0.212; k = 3 chosen on interpretability, as silhouette alone favoured k = 2): upright standing and locomotion (55.9% of frames), squatting and kneeling (23.1%), and deep stooping (21.0%). Frame-level OWAS coding with the load digit fixed at class 1 placed 52.8% of frames in AC2 and 13.8% in AC3 (29.5% in AC3 under an upper-bound load-class-2 scenario), with high-risk exposure ranging from 0.0% to 24.2% across recordings; no frame reached AC4, which in the OWAS matrix requires a twisted back or a load above 20 kg. Anomaly density and AC exposure rankings diverged, indicating complementary information.

Originality and Practical Implications: To our knowledge this is among the first leakage-audited, label-free pipelines to turn raw motor-manual forestry video into per-recording OWAS exposure profiles, offering a scalable, non-intrusive complement to manual audits and a template for continuous digital ergonomic surveillance in outdoor, high-risk occupations. Findings are exploratory and are not validated against motion-capture or expert-coded OWAS labels.

Clinical trial number: not applicable

Keywords: Forestry ergonomics; OWAS; motor-manual operations; pose estimation; unsupervised learning; autoencoder; anomaly detection; occupational safety

1. Introduction

1.1. Background

Forestry is consistently ranked among the most hazardous occupational sectors worldwide, and logging workers record some of the highest work-related fatality rates of any occupational group in the United States [1, 2]. While acute traumatic injury dominates fatality statistics, the cumulative postural loading generated by motor-manual work, namely felling, delimbing, and cross-cutting standing timber with a handheld chainsaw rather than a fully mechanised harvester, is a recognised and comparatively under-monitored driver of work-related musculoskeletal disorders (WMSDs) among surviving operators [3, 4]. Motor-manual technique remains widespread in smallholder forestry, on steep or inaccessible terrain, and across much of the developing world, despite decades of progress in mechanised cut-to-length systems.

Chainsaw work compels the operator into a narrow, tool-dictated envelope of postures: a forward-flexed and often laterally rotated trunk to maintain visual contact with the cut, asymmetric arm elevation to manoeuvre the guide bar around branches and stems, and sustained squatting, kneeling, or half-kneeling leg configurations to delimb material close to the ground. These are the classical biomechanical antecedents of low back, shoulder, and knee WMSDs, and because they are sustained across entire working shifts, their cumulative rather than instantaneous magnitude ultimately determines musculoskeletal risk.

Ergonomic science has responded to such hazards with a family of observational posture-assessment tools, of which OWAS is among the oldest and most field-practical [5]. OWAS was designed for use by non-specialist observers directly on the shop floor. It encodes a worker's instantaneous posture as a four-digit code describing the back, arms, legs, and handled load, and maps that code through a fixed lookup table onto one of four Action Categories (AC1–AC4), ranging from 'no corrective action needed' to 'corrective action needed immediately'. It has since become one of the most widely used postural assessment methods in forest operations ergonomics, together with a continuing discussion of its advantages and limitations in that setting [3, 6]. Alongside OWAS, the Rapid Upper Limb Assessment (RULA) and Rapid Entire Body Assessment (REBA) provide upper-limb- or whole-body-weighted scoring systems that are frequently

cross-compared with OWAS [7, 8]. Comparisons of OWAS, RULA, REBA, and related tools in farm work show that the tools can yield different risk estimates for the same task [9], and a forestry-specific comparison on manual wood-chipper feeding reported that RULA was the more precautionary of the two methods examined for the tasks studied [10].

The persistent limitation shared by all observational methods, OWAS included, is that they rely on discrete, human-coded sampling of postures by a trained analyst watching the worker in person or on video. This is defensible for short task studies but scales poorly to full-shift, multi-worker, multi-site surveillance, where the annotation burden quickly becomes unsustainable. Early attempts to automate observational scoring with depth sensors illustrate how long this problem has been recognised [11], and the problem has motivated a rapidly growing computer-vision literature that scores posture directly from video using markerless pose estimation frameworks such as OpenPose, AlphaPose/RMPE, and MediaPipe [12–14].

Vision-based classifiers built on skeletal keypoints and view-invariant features have been developed for construction hazard prevention [15]. Incremental deep-network classifiers have been trained for on-site ergonomic risk assessment [16], variational deep architectures have been proposed for automatic ergonomic risk assessment [17], and webcam-based deep-learning pipelines have been demonstrated outside the laboratory [18]. Within forestry, transfer-learning classifiers applied to more than 5,000 images of motor-manual cross-cutting have been used to automate OWAS classification [4]. A subsequent reliability study found that such a deep-learning model assessed posture 19 to 53 times faster than human raters, while agreement between individual human raters and the model ranged from poor to fair (Cohen's kappa −0.03 to 0.34), and rose to fair-to-moderate when all raters were pooled with the model [6]. Self-occluded postures relevant to outdoor field conditions have been addressed with dedicated deep architectures [19]. In parallel, automatic feature selection has been shown to support activity recognition for ergonomics assessment of industrial tasks [20], and wearable inertial and insole sensors provide a complementary, non-visual channel for recognising awkward postures on occluded or outdoor worksites [21–23].

Two weaknesses nonetheless characterise this literature. First, most reported systems are supervised or rule-based classifiers that require manually annotated posture labels or tuned decision thresholds, and even real-time vision-based postural-risk methods are typically evaluated on defined task sets [24]. Such approaches do not scale readily to the diversity of postures produced by unscripted field labour. Second, few studies have been conducted in outdoor, occlusion-heavy environments such as an active forest stand, where dense vegetation, exhaust, flying sawdust, and changing illumination degrade pose-tracking confidence in ways that laboratory or factory validation does not anticipate.

A largely independent line of research has shown that reconstruction-based unsupervised deep learning, and specifically the deep autoencoder, is well suited to detecting anomalous patterns without labelled abnormal examples. An autoencoder trained on normative data learns a compressed latent manifold of typical configurations, and inputs that deviate from it yield disproportionately high reconstruction error [25, 26]. This reconstruction-error paradigm, alongside classical outlier detectors such as the Isolation Forest and the One-Class Support Vector Machine, is well established in the general anomaly-detection literature [27, 28]. To our knowledge, however, it has not been combined with a standardised occupational coding scheme such as OWAS to produce a continuous, label-free, video-derived exposure metric for motor-manual forestry.

The gap is compounded by a methodological concern in applied machine learning: unaudited temporal or subject-level data leakage inflates reported performance across many scientific disciplines, motivating explicit leakage-prevention protocols for any pipeline that, like ours, partitions highly correlated video frames [29]. Meanwhile, the movement toward human-centric, AI-augmented Industry 5.0 work systems has articulated a need for scalable, non-intrusive digital tools that treat continuous worker well-being monitoring as a design objective alongside productivity, rather than as a periodic, resource-intensive audit [30].

The rationale for an unsupervised approach is practical as much as statistical. Supervised posture classifiers inherit the cost and the subjectivity of the labels used to train them: someone must code thousands of images, and the resulting model can only reproduce the coding scheme, and the coder biases, that it was taught. Human raters applying OWAS to the same forestry imagery do not always agree with one another or with a trained model, with agreement ranging from poor to moderate depending on how raters are combined [6]. An approach that first discovers the structure of the posture data itself, and only then attaches an interpretive coding scheme, does not remove subjectivity from the translation step, but it makes the dependence explicit and keeps the expensive human labour at the stage where it is most informative, namely reviewing the frames the model flags.

The setting also matters. Motor-manual forestry footage is captured under conditions that routinely violate the assumptions of laboratory pose-estimation benchmarks: the camera is often hand-held or chest-mounted, the worker is partially hidden by stems and undergrowth, and lighting changes with canopy and weather. A pipeline intended for operational surveillance must therefore treat tracking failure as a normal event, filter degraded frames conservatively, and be evaluated in a way that does not reward memorising the idiosyncrasies of a particular recording. These requirements shaped both the visibility filter and the video-level validation split used in this study.

1.2. Research Aim, Objectives and Questions

The overarching aim of this study was to design, implement, and validate a fully unsupervised, video-based computational pipeline capable of quantifying OWAS ergonomic risk exposure in authentic motor-manual forestry footage, without requiring manually labelled ground-truth postures at any stage of model development.

Three objectives followed. The first was methodological: to build a leakage-audited, reproducible pipeline that converts raw field video into biomechanical joint-angle features, partitions them strictly by video

source, and trains complementary classical and deep unsupervised models. The second was diagnostic: to interrogate the stability, sensitivity, and concordance of these models through architectural, feature-level, and cross-detector ablation studies, and to quantify agreement between independent paradigms with Cohen's kappa so that only cross-validated anomalies would be retained. The third was translational: to map unsupervised cluster centroids and anomaly flags onto the standardised four-digit OWAS scheme and its Action Categories, producing per-recording and corpus-wide exposure profiles, and an automatically highlighted verification video, that practitioners can interpret without machine-learning expertise.

Three research questions guided the work. RQ1: Can a compact deep autoencoder, trained only on biomechanical features from unscripted, occlusion-heavy field video, learn a stable low-dimensional representation of normative motor-manual posture that generalises across video sources under a strict group-based split? RQ2: To what extent do independent unsupervised paradigms (tree-based isolation, kernel-based one-class classification, and reconstruction-based deep learning) converge on the same frames as postural outliers? RQ3: Can unsupervised posture clusters and consensus anomalies be translated into the OWAS Action Category framework to yield a quantitative exposure metric that is plausible for motor-manual forestry?

The study makes four contributions. First, it demonstrates a complete label-free route from raw forestry video to OWAS Action Category exposure profiles. Second, it documents an explicit, reportable leakage-prevention protocol (video-level grouping, training-fold-only scaling, and training-derived thresholds) for correlated video-frame data. Third, it shows through three ablation studies how the main design choices of the deep and classical branches behave and how much the different detector families agree. Fourth, it shows empirically that an anomaly-density measure and a standardised posture-category exposure measure rank the same recordings differently, which has implications for how digital ergonomic surveillance should be reported. The remainder of the paper is organised as follows. Section 2 develops the theoretical and conceptual framework; Section 3 describes the materials and methods; Section 4 reports the results; Section 5 discusses interpretation, implications, ethical considerations, and limitations; and Section 6 concludes.

2. Literature Review

2.1. Theoretical Framework

This study draws on three complementary theoretical traditions: observational ergonomics, biomechanical exposure theory, and statistical representation learning. Observational ergonomics, exemplified by OWAS, treats posture as a discrete, classifiable state rather than a continuous signal [5]. Its four-digit code reduces an effectively infinite space of body configurations to a tractable categorical system that a human observer can apply in the field, and its Action Category table translates each code into an explicit intervention priority, from AC1 (no action needed) to AC4 (act immediately). RULA and REBA extend the tradition with segment-weighted scoring that also considers muscle activity and force; benchmarking against OWAS in agricultural and forestry contexts shows that the tools emphasise different body regions and can therefore yield different risk estimates for the same task [7–10].

Biomechanical exposure theory supplies the causal mechanism linking posture to harm: sustained or repeated deviation from neutral joint angles raises passive tissue strain and compressive spinal loading, and it is the cumulative, time-integrated exposure to non-neutral postures, rather than any single instantaneous posture, that drives WMSD risk over a working career. This is why expressing the percentage of analysed frames in each Action Category is more meaningful than a single point-in-time score, and it justifies the exposure-profiling stage of the present pipeline (Section 3.10). It also motivated the seven kinematic features (trunk flexion, trunk lateral tilt, bilateral arm elevation, bilateral knee flexion, and hip drop ratio), which span the anatomical dimensions that OWAS, REBA, and RULA treat as primary drivers of postural load.

Representation-learning theory provides the statistical foundation of the modelling core. An autoencoder is trained to reconstruct its own input through a compressed latent bottleneck and is thereby forced to discover

a lower-dimensional manifold capturing the regularities of the training distribution [25, 31]. When that distribution is dominated by task-typical postures, as in continuous unscripted field recording where most frames depict routine work, the learned manifold approximates a model of normality, and reconstruction error becomes a label-free proxy for postural deviation [26, 32, 33]. Training is stabilised by batch normalisation [34], a leaky rectified-linear activation that avoids 'dying units' [35], and decoupled weight decay [36], which builds on the adaptive optimiser of Kingma and Ba [37].

A complementary shallow-learning tradition treats anomaly detection as geometric or density-based isolation. The Isolation Forest exploits the fact that anomalous points need, on average, fewer random recursive partitions to be isolated than points embedded in a dense cluster [38], extending the ensemble-tree paradigm of Random Forests [39]. The One-Class Support Vector Machine instead learns a smooth boundary enclosing the bulk of the training distribution in a kernel-induced feature space [40]. Reviews treat these statistical, isolation-based, and reconstruction-based families as complementary approaches to one problem [27, 28]. Because each detector formalises 'abnormality' through a different mechanism, agreement between two of them is stronger evidence of genuine deviation than either output alone, an inferential logic formalised by Cohen's [41] chance-corrected kappa and interpreted against the qualitative benchmarks of Landis and Koch [42].

Finally, clustering theory and its validity indices supply the apparatus for discovering latent postural archetypes without supervision. K-Means partitions observations to minimise within-cluster variance [43, 44], with stability improved by principled seeding [45]. Because K-Means works best on decorrelated, variance-ranked axes it is often paired with Principal Component Analysis [46–48], a pairing with a formal connection: the principal components provide a continuous relaxation of the cluster-membership indicators sought by K-Means [49]. With no ground-truth posture labels, we relied on internal indices: the silhouette coefficient [50, 51], which has been used to select and interpret movement-pattern clusters in other biomechanical domains such as shoulder range of motion [52]; the Davies–Bouldin index [53]; and the Calinski–Harabasz index [54]. Chance-corrected partition-agreement statistics [55] offer a further basis for

comparing alternative cluster solutions, and general treatments of these methods are given by Hastie, Tibshirani, and Friedman [56].

2.2. Conceptual Framework

Building on these foundations, we developed the conceptual framework shown in Figure 1, which organises the study into five interlocking layers: data acquisition, kinematic representation, dual-paradigm unsupervised modelling, cross-model consensus and validation, and ergonomic translation. Figure 1 also shows, in its top band, the three theoretical traditions introduced above, with dashed arrows indicating which stage of the pipeline each one informs. The framework makes explicit the logic connecting each tradition to an auditable stage of the pipeline, and it is intended to generalise beyond forestry to any occupation in which continuous video is available but manual posture annotation is infeasible at scale.

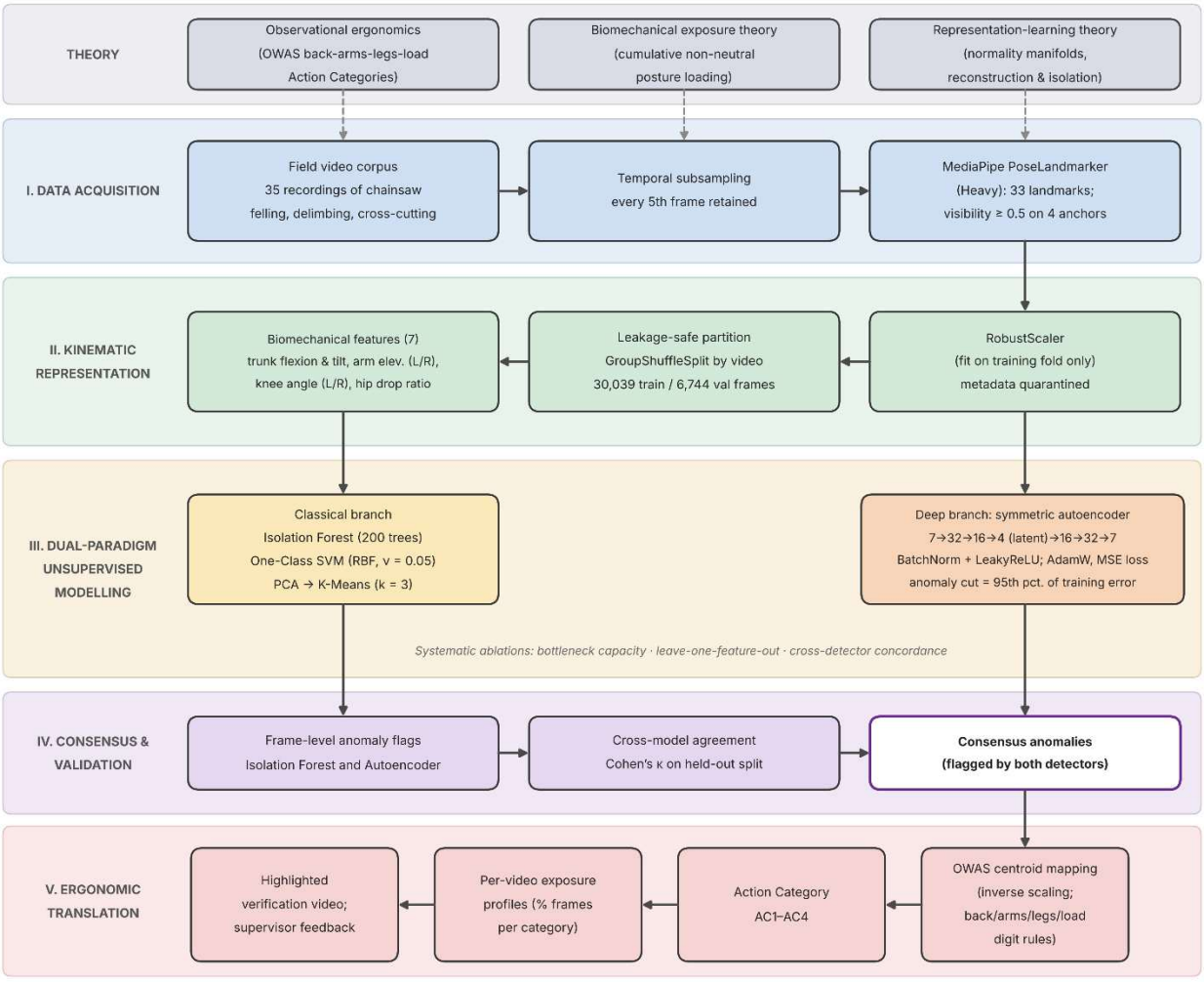


Figure 1. Conceptual framework for unsupervised, video-based OWAS ergonomic risk quantification in motor-manual forestry. Layers I–V run from data acquisition to ergonomic translation; the top band shows the theoretical traditions informing each stage; dashed arrows indicate theory-to-stage links.

In the Data Acquisition Layer (Figure 1, Layer I), unscripted video of motor-manual chainsaw operations is temporally subsampled and passed through a markerless pose estimator that yields three-dimensional landmarks [14], consistent with the observational-ergonomics premise that posture can be characterised adequately from external visual inspection. A visibility filter on four bilateral anchors guards against degraded skeletons. The Kinematic Representation Layer (Layer II) converts retained landmarks into the seven anatomically grounded features motivated by biomechanical exposure theory, enforces a video-level,

leakage-safe train/validation partition, and fits the feature scaler on the training fold only, operationalising the leakage-prevention principle of Kapoor and Narayanan [29].

The Dual-Paradigm Unsupervised Modelling Layer (Layer III) instantiates the representation-learning and shallow-isolation traditions in parallel rather than in competition. A classical branch (Isolation Forest, One-Class SVM, and PCA-reduced K-Means) is trained alongside a deep branch consisting of the symmetric autoencoder described in Section 3.6. Ablation studies of bottleneck capacity, feature sensitivity, and cross-detector correlation (Section 3.7) sit at the boundary between this layer and the next, because their purpose is to establish how much confidence the consensus layer should place in any individual model.

The Cross-Model Consensus and Validation Layer (Layer IV) operationalises chance-corrected agreement by admitting a frame to the high-confidence anomaly set only if both the Isolation Forest and the autoencoder flag it, converting two individually noisy detectors into a more conservative signal. Finally, the Ergonomic Translation Layer (Layer V) closes the loop with observational ergonomics: cluster centroids are inverse-transformed into physical units and mapped through the OWAS decision rules [5] to four-digit codes and Action Categories, which are aggregated into per-recording exposure profiles and rendered as annotated verification video, the two outputs intended for practitioners who need not interact with the underlying models.

The architecture is deliberately modular. The pose estimator, the unsupervised detectors, or the coding scheme can in principle be substituted (for example, REBA or RULA rules could replace the OWAS lookup in Layer V) without invalidating the framework, positioning it as a template for label-free ergonomic surveillance consistent with calls for scalable, human-centric monitoring in Industry 5.0 work systems [30].

2.3. Research Gap and Guiding Propositions

Taken together, the literature reviewed above supports a clear, bounded research gap. Observational ergonomics provides a trusted, interpretable endpoint (OWAS Action Categories) but not a scalable way of

reaching it; computer-vision systems provide scalability but, in the studies reviewed here, generally depend on labelled data or on rule sets validated for defined task sets; and unsupervised anomaly detection provides label-free deviation scoring but has not been tied to a standardised ergonomic coding scheme in motor-manual forestry. The gap is therefore not the absence of any one component but the absence of a validated, leakage-audited chain linking them, together with evidence on how the components behave when linked.

To keep the empirical work disciplined, we express the framework in Figure 1 as four propositions that the results in Section 4 can support or contradict. Proposition 1 (stability): a compact autoencoder trained on seven biomechanical features will converge to a stable latent representation of normative posture that holds on videos unseen during training. Proposition 2 (triangulation): detectors built on different mathematical assumptions will agree on anomaly status beyond chance but only partially, so that cross-paradigm consensus yields a more conservative anomaly set than any single detector. Proposition 3 (structure): the posture data contain a small number of recurrent archetypes that correspond to the task repertoire of motor-manual felling, delimbing, and cross-cutting, and that are interpretable through physical-unit centroids. Proposition 4 (complementarity): an anomaly-density measure and an OWAS category-exposure measure capture different aspects of postural risk and therefore will not rank recordings identically.

These propositions are deliberately modest. They do not assert that the pipeline reproduces expert OWAS coding, that the discovered archetypes are the only valid partition of the data, or that anomalous frames are necessarily hazardous. They assert only that the components behave as the underlying theory predicts when assembled, which is the minimum condition for the framework to be useful as a screening layer. The results in Section 4 are evaluated against these propositions in the Discussion.

3. Materials and Methods

3.1 Video Acquisition and Data Preprocessing

We collected video recordings of motor-manual forest operations under natural field conditions at six Forest Directorates in Romania. Forestry workers performed tree felling, delimbing, and cross-cutting operations using standard professional chainsaws. Verbal informed consent was obtained from all participants, which outlined the research objectives, the voluntary nature of participation, and respondents' right to withdraw at any time before submission. Ethical clearance for this study was reviewed and approved by the Ethics Committee of Transilvania University of Brașov (UNITBV) on 15 July 2025. As the study involved a minimal-risk remote filming of professional chainsaw operators, institutional clearance was granted without an assigned reference protocol number. We stored all raw video recordings in a local directory containing standard digital video formats, including MP4, AVI, and MOV container files across thirty-five distinct operational video recordings. Postures during motor-manual forest operations transition across seconds rather than tiny fractions of a second. We therefore downsampled each video stream by extracting every fifth frame. This temporal subsampling eliminated redundant kinematic data between adjacent frames and minimized downstream computational overhead while preserving rapid postural transitions during operational cycles.

3.2 Pose Estimation and Quality Filtering

We tracked body kinematics using the Google MediaPipe Tasks computer vision framework [14]. Specifically, we deployed the heavy variant of the PoseLandmarker model architecture, which infers thirty-three anatomical landmarks per subject in three-dimensional camera space. Outdoor forest environments present severe visual obstacles, including tree trunks, dense underbrush, flying sawdust, and chainsaw exhaust smoke, all of which frequently obstruct human body segments. To maintain data integrity, we enforced a landmark tracking confidence threshold of 0.5 across four critical bilateral anatomical anchors:

the left shoulder, right shoulder, left hip, and right hip. The automated ingestion pipeline discarded any video frame where the visibility confidence of any of these four reference landmarks fell below 0.5. This strict filter prevented degraded skeletal estimations from entering downstream models.

3.3 Biomechanical Feature Engineering

We converted the spatial joint coordinates of every accepted frame into seven continuous biomechanical variables representing the primary kinematic chains assessed by observational ergonomics. Planar joint angles were computed using the vector dot product. For three consecutive joint coordinates **a**, **b**, and **c**

forming a joint vertex at coordinate **b**, the formula computes the enclosed planar angle:

$$\theta = \arccos\left(\frac{(\mathbf{a}-\mathbf{b})\cdot(\mathbf{c}-\mathbf{b})}{\|\mathbf{a}-\mathbf{b}\|\|\mathbf{c}-\mathbf{b}\|}\right) \text{ (Equation 1)}$$

All computed cosine quotients were clamped to the numerical interval $[-, 1.01.0]$ to prevent floating-point domain errors. The first variable measured trunk flexion as the planar angle between the torso vector and an upward vertical reference vector. The second variable quantified lateral trunk tilt using the absolute difference between the shoulder slope and hip slope, capturing spinal lateral bending and torsion. The third and fourth variables measured left and right arm elevation angles relative to the torso vector. The fifth and sixth variables captured left and right knee flexion angles from hip, knee, and ankle vertex coordinates.

The seventh variable computed the hip drop ratio by dividing vertical lower-body height by vertical torso height. In natural field video tracking, extreme forward bending or momentary torso foreshortening can cause vertical torso height to approach zero, resulting in division-by-zero numerical explosions. To prevent artificial distortion, we clamped the hip drop ratio to biologically plausible bounds between 0.10 and 2.50, and imputed any residual indeterminate values using the empirical feature median. This continuous bounded ratio reliably separated upright standing from stooping, kneeling, and deep squatting postures without introducing mathematical artifacts into subsequent loss functions.

3.4 Data Leakage Prevention and Partitioning Protocol

Because adjacent video frames exhibit high temporal autocorrelation, standard randomized frame-level partitioning leaks information between contiguous time steps and artificially inflates validation metrics. To eliminate data leakage, we enforced a four-stage audit protocol prior to model ingestion.

First, we implemented a group-based splitting protocol partitioned strictly by video source identity (GroupShuffleSplit). All frames belonging to any single video recording were allocated exclusively to either the training partition (eighty percent of the total dataset, comprising 30,039 frames) or the holdout validation partition (twenty percent, comprising 6,744 frames), guaranteeing zero temporal overlap across partitions. For single-recording analyses, a chronological block split was maintained to preserve temporal isolation.

Second, we strictly quarantined all non-kinematic tracking metadata, including video identifiers, frame indices, and monotonic time stamps (timestamp_sec), ensuring that models evaluated purely biomechanical joint kinematics rather than temporal progression.

Third, to eliminate preprocessing leakage, we standardized all features using a RobustScaler based on the median and interquartile range to resist sensor noise, fitting the scaling parameters strictly on the training partition and applying them post-hoc to transform validation and full-dataset frames.

Fourth, we eliminated threshold calibration leakage by deriving all downstream anomaly decision thresholds, including the ninety-fifth percentile reconstruction cutoff, exclusively from the empirical distribution of the training partition rather than from the validation or testing sets.

3.5 Unsupervised Machine Learning Framework

We trained classical unsupervised anomaly detection and clustering models on the leak-free, scaled training representations. We initialized an Isolation Forest ensemble of two hundred isolation trees with a contamination hyperparameter of five percent [38]. The algorithm recursively partitions kinematic features

by randomly selecting split thresholds, assigning anomaly scores based on the path length required to isolate a given posture. In parallel, we fitted a One-Class Support Vector Machine utilizing a Radial Basis Function kernel with a five percent outlier boundary to establish a smooth non-linear decision boundary around standard postures [40]. To discover postural archetypes, we applied K-Means clustering with k-means++ initialization (ten initializations, fixed random seed) to the seven-dimensional scaled feature space of the complete corpus [45]. Because K-Means uses no labels, fitting it on all frames does not leak label information, but cluster assignments for validation frames were not predicted by a model fitted on the training partition alone. Principal Component Analysis with two components was used to visualize the posture space and as a linear reconstruction-error baseline detector. Candidate cluster counts from two to ten were compared post hoc with the validity indices described in Section 3.9, and k = 3 was retained on grounds of interpretability and cluster balance rather than as an index optimum (Section 4.4).k

3.6 Unsupervised Deep Learning Architecture

To capture complex non-linear postural representations without manual supervision, we constructed a deep symmetric autoencoder in PyTorch. The model compressed the seven normalized biomechanical features into a compact latent bottleneck before reconstructing the input vector. The encoder projected the seven inputs to thirty-two hidden units, followed by one-dimensional batch normalization and a LeakyReLU activation function with a negative slope of 0.2. A second fully connected layer compressed the representation to sixteen hidden units with batch normalization and LeakyReLU activation, after which a final linear projection produced a four-dimensional latent bottleneck representation.

The decoder symmetrically mirrored this architecture, expanding the four-dimensional latent representation through fully connected layers of sixteen and thirty-two hidden units with batch normalization and LeakyReLU activations before reconstructing the seven original biomechanical features. We trained the network for forty epochs using a mini-batch size of sixty-four samples and the AdamW optimizer with a

learning rate of 0.001 and a weight decay parameter of 0.0001 [36]. The objective function minimized the Mean Squared Error between original and reconstructed feature vectors:

$$\text{MSE} = \frac{1}{D}\sum_{j=1}^{D}(x_j - \hat{x}_j)^2 \quad \text{(Equation 2)}$$

Frames exhibiting high reconstruction errors indicated biomechanical configurations that deviated significantly from standard movement cycles. We established an anomaly decision threshold at the ninety-fifth percentile of the training reconstruction error distribution, classifying frames exceeding this cutoff as deep learning postural anomalies.

3.7 Systematic Ablation Studies

To verify model stability and justify architectural choices, we conducted three systematic ablation studies. The first ablation investigated latent bottleneck capacity and normalization by systematically training autoencoder variants with bottleneck dimensions of two, three, four, and six units, alongside a companion architecture omitting batch normalization layers. Tracking validation loss convergence across training epochs enabled us to identify the optimal bottleneck capacity that prevented identity memorization while preserving reconstruction fidelity. Each configuration was trained for 30 epochs. The original analysis was a single PyTorch run; to assess seed sensitivity it was replicated with five random seeds using an independent NumPy implementation of the same architecture and optimizer (Section 3.10). The replication also measured how well each bottleneck's reconstruction error separated validation frames flagged by the Isolation Forest from unflagged frames (error ratio and area under the ROC curve), which provides a direct test of the identity-memorization concern.

The second ablation assessed feature sensitivity through an iterative leave-one-feature-out protocol. We repeatedly retrained the autoencoder while omitting each biomechanical variable one at a time and measured the resulting delta in validation reconstruction error relative to the baseline model. This isolated the kinematic influence of each joint angle and identified which body segments most strongly governed

stability representations. In the original run the baseline model was trained for 30 epochs and each reduced model for 25 epochs, and the deltas compared overall reconstruction error, which is averaged over a different number of inputs. The replication therefore trained all models for 30 epochs and compared each reduced model with the baseline model on the retained features only (matched-feature delta).

The third ablation benchmarked anomaly detector concordance by comparing frame-level anomaly predictions across Isolation Forest, One-Class Support Vector Machine, PCA reconstruction error, and the deep autoencoder. Computing pairwise correlation matrices across the held-out validation split established the degree of consensus among diverse unsupervised detection paradigms.

3.8 OWAS Coding Protocol

We applied the OWAS decision rules [5] at two levels. The OWAS system defines postures using four digits representing back, arms, legs, and load. The rules below were applied (i) to every frame, which forms the basis of the exposure profiles, and (ii) to the cluster centroids, which were transformed back into physical degree and ratio values using inverse scaling, to interpret the archetypes.

Trunk flexion below twenty degrees mapped to a straight back in digit 1, whereas trunk flexion at or above twenty degrees mapped to a bent back in digit 2. Lateral tilt values exceeding fifteen degrees mapped to a twisted or sideways bent back in digits 3 or 4. Arm elevations below eighty-five degrees mapped to both arms below shoulder level in digit 1, single arm elevation above eighty-five degrees mapped to one arm elevated in digit 2, and bilateral arm elevation above eighty-five degrees mapped to both arms elevated in digit 3.

Average knee flexion angles and hip drop ratios determined leg postures across digits 1 through 7. Straight knee angles exceeding one hundred fifty degrees with normal hip drop ratios indicated standing in digit 2. Knee flexion below one hundred twenty degrees with uneven extension mapped to one knee bent in digit

5. Bilateral knee flexion below one hundred thirty degrees mapped to squatting in digit 4. Low hip drop ratios combined with knee flexion indicated kneeling or sitting in digits 1 or 6.

The load digit was fixed at class 1 (handled load below ten kilograms) for all frames, because the mass of a professional chainsaw (approximately six to nine kilograms) lies below that limit and the pipeline contains no tool-engagement classifier with which to vary the digit by task phase. Because load class 2 (ten to twenty kilograms) could be argued for active bucking and felling when cutting forces are considered, we also computed an upper-bound sensitivity scenario in which every frame is assigned load class 2; true exposure under such an argument would lie between the two scenarios. Each four-digit profile indexed an OWAS action category ranging from 1 to 4, designating normal posture, slightly harmful posture, distinctly harmful posture, or extremely harmful posture requiring immediate intervention.

3.9 Diagnostic Evaluation and Validation Metrics

Because continuous operational video lacks frame-by-frame manual labels, model performance was evaluated through intrinsic clustering metrics and inter-model consensus diagnostics. We computed the Silhouette score to quantify cluster cohesion and separation on a range from negative one to positive one [50]. We calculated the Davies-Bouldin Index to assess cluster compactness relative to separation, where lower values designate superior partitioning [53]. We determined the Calinski-Harabasz Index to quantify the ratio of between-cluster variance to within-cluster dispersion, with higher scores reflecting denser clustering [54].

Autoencoder convergence was tracked using continuous training and validation loss curves across epochs to confirm the absence of overfitting. We constructed a cross-model contingency matrix comparing Isolation Forest anomaly classifications against autoencoder anomaly predictions on the held-out validation split, quantifying concordance using Cohen's Kappa coefficient [41]. High inter-model agreement identified high-confidence consensus anomalies representing genuine biomechanical deviations.

3.10 Exposure Profiling, Verification Video, and Software

Exposure profiling followed directly from the OWAS coding. The decision rules of Section 3.8 were applied at the level of individual frames, so that every retained frame received back, arm, leg, and load digits and an Action Category from the full four-dimensional OWAS lookup matrix; the physical-unit cluster centroids were reported alongside to characterise the archetypes (Table 5). In the implementation reported here the load digit was fixed at class 1 for all frames, reflecting the mass of a professional chainsaw (below 10 kg). For each recording we then computed the percentage of frames in each Action Category, the combined AC3 + AC4 ('high-risk') percentage, the number of consensus anomalies, and the anomaly density (consensus anomalies divided by analysed frames). Corpus-wide percentages were computed across all frames.

For verification, an annotated video was rendered for a selected recording in which a status panel displays the current OWAS code and Action Category and the frame border and panel turn red when a frame is a consensus anomaly or reaches AC4 (orange for AC3, green otherwise); annotations were updated on the sampled frames and held between them. Analyses were run in Python using scikit-learn for the Isolation Forest, One-Class SVM, PCA, K-Means, scaling, and validity indices [57], and PyTorch for the autoencoder [58].

Three post hoc analyses were added to resolve choices that the primary pipeline had fixed. First, K-Means was re-run for k = 2 to 10 on the scaled corpus (ten initializations, seed 42) and compared by silhouette (10,000-frame subsample), Davies–Bouldin, and Calinski–Harabasz indices. Second, OWAS codes were computed for the physical-unit centroids of the retained clusters. Third, the OWAS lookup was repeated with the load digit set to class 2 for all frames to bound the effect of the load assumption. In addition, Ablations 1 and 2 were replicated with five seeds (0–4) and 30 epochs for every model, using a NumPy implementation of the autoencoder (seven-input, 32–16–bottleneck–16–32–seven layers, batch normalization, LeakyReLU with slope 0.2, AdamW with learning rate 0.001 and weight decay 0.0001, batch size 64, mean squared error), whose gradients were verified against finite differences. PyTorch was

not available for this replication, so its absolute values differ somewhat from the original run; Tables 2 and 3 report both.

4. Results

4.1 Data Partition and Leakage Audit

The pipeline processed 36,783 pose-estimated frames from 35 motor-manual forestry recordings. After the leakage-safe GroupShuffleSplit partition, 30,039 frames (81.7%) formed the training set and 6,744 frames (18.3%) the held-out validation set, with no video shared between partitions, as confirmed by the automated leakage audit; the RobustScaler and all anomaly thresholds were fitted on the training partition only [29]. Table 1 consolidates the sample characteristics, partition sizes, and headline statistics that are referred to throughout this section.

Table 1. Sample characteristics, data partition, and headline validation statistics.

Characteristic	Value
Field recordings analysed	35
Frame sampling interval	Every 5th frame
Accepted frames (landmark visibility ≥ 0.5 on 4 anchors)	36,783
Training partition	30,039 frames (81.7%)
Held-out validation partition	6,744 frames (18.3%)
Video overlap between partitions	None (leakage audit)
Biomechanical features	7

Characteristic	Value
Autoencoder validation MSE (4-unit bottleneck)	0.0886 (original run); 0.0819 ± 0.0022 (5 seeds)
Cohen's κ, Isolation Forest vs autoencoder (validation split)	0.431
Consensus anomalies (full corpus)	777 (2.11%)
K-Means clusters retained	3 (judgement-based; Table 4)
Silhouette / Davies–Bouldin / Calinski–Harabasz (k = 3)	0.212 / 1.756 / 7,266.4

Note. MSE = mean squared error. Silhouette computed on a random 10,000-frame subsample.

4.2 Ablation Studies

Figure 2 and Table 2 report the first ablation study, which varied the autoencoder's latent bottleneck size and the presence of batch normalisation. In the original single PyTorch run, validation mean squared error (MSE) fell as bottleneck capacity grew, from 0.2434 with two units to 0.1550 with three, 0.0886 with four (the production configuration), and 0.0174 with six, while the four-unit network without batch normalisation reached 0.0798 (its curve in Figure 2 lay below the batch-normalised curve throughout training). Because a single run cannot indicate seed sensitivity, the ablation was replicated with five seeds (Section 3.10). The replication reproduced the original values closely for most configurations (mean validation MSE 0.2456 ± 0.0141 for two units, 0.1376 ± 0.0073 for three, 0.0819 ± 0.0022 for four, and 0.0188 ± 0.0014 for six), and between-seed variation was small relative to the differences between bottleneck sizes, so the decrease in error with capacity is robust. For batch normalisation the two analyses disagreed: the single run gave lower error without it, whereas the replication gave slightly lower error with it (0.0819 ± 0.0022 versus 0.0847 ± 0.0036). The difference is of the order of one standard deviation, so no firm conclusion about the effect of batch normalisation on final error can be drawn.

Table 2. Ablation 1: validation reconstruction error and anomaly separation by autoencoder bottleneck size and normalisation (original single PyTorch run and five-seed NumPy replication).

Configuration	Single run MSE	5-seed MSE (mean ± SD)	Error ratio, flagged / unflagged	AUC vs Isolation Forest flags
2 units, batch normalisation	0.2434	0.2456 ± 0.0141	3.70 ± 0.81	0.875 ± 0.018
3 units, batch normalisation	0.1550	0.1376 ± 0.0073	3.21 ± 0.28	0.864 ± 0.030
4 units, batch normalisation (production model)	0.0886	0.0819 ± 0.0022	3.72 ± 0.09	0.878 ± 0.012
6 units, batch normalisation	0.0174	0.0188 ± 0.0014	3.27 ± 0.21	0.822 ± 0.020
4 units, no batch normalisation	0.0798	0.0847 ± 0.0036	3.37 ± 0.72	0.821 ± 0.024

Note. Error ratio and AUC are computed on the validation split with Isolation Forest flags as the reference; they measure concordance with that detector, not accuracy against true anomalies. Single-run values come from the original PyTorch analysis; seed-based values from the NumPy replication.

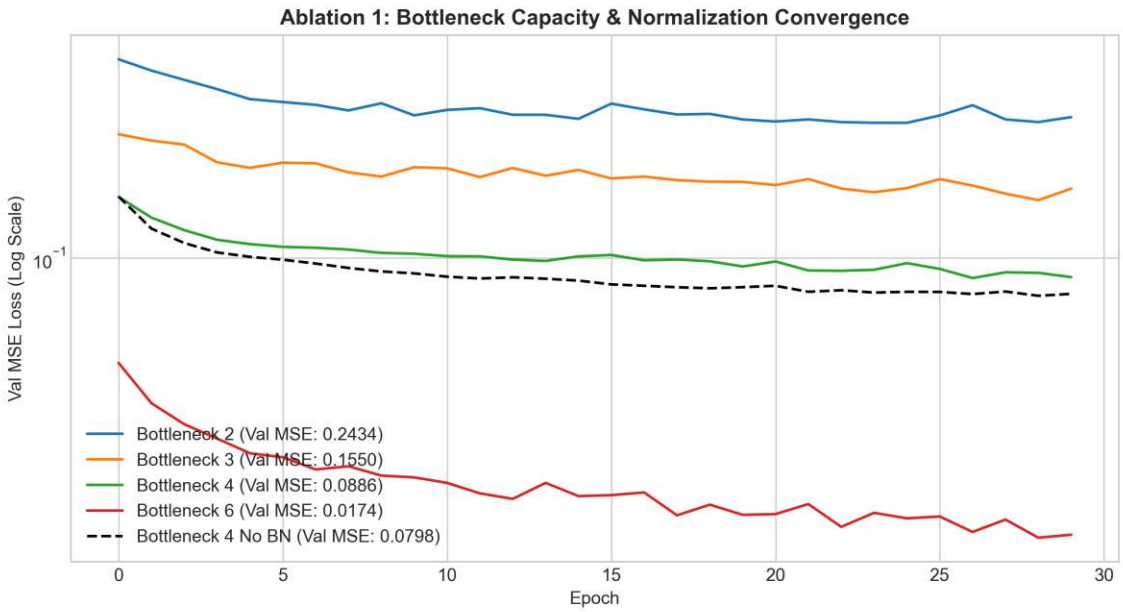


Figure 2. Ablation 1 (original single run): validation MSE (log scale) across 30 epochs for each bottleneck configuration, with and without batch normalisation (BN).

The replication also tested directly whether wider bottlenecks erode the separation of anomalous from normal frames, using Isolation Forest flags on the validation split as the reference. Both the ratio of mean reconstruction error in flagged to unflagged frames and the ROC area under the curve were highest at four units (ratio 3.72 ± 0.09; AUC 0.878 ± 0.012). Two and three units gave similar AUCs (0.875 ± 0.018 and 0.864 ± 0.030) but much higher overall error, and six units gave the lowest separation (ratio 3.27 ± 0.21; AUC 0.822 ± 0.020), even though its overall reconstruction error was lowest. This supports retaining four units: the six-unit network reconstructs better overall but separates flagged from unflagged frames less well, consistent with a bottleneck that is wide enough to start reconstructing atypical frames too [25, 26]. Because the reference flags come from another detector, this measures concordance with the Isolation Forest rather than accuracy against true anomalies.

Figure 3 and Table 3 report the second ablation, a leave-one-feature-out analysis. In the original single run (baseline MSE 0.0772 at 30 epochs; reduced models trained for 25 epochs), every omission lowered overall

validation MSE, by 0.0159 to 0.0267, most for right arm elevation, hip drop ratio, and right knee angle. That design had two limitations: the reduced models were trained for fewer epochs than the baseline, and overall MSE averages over different numbers of inputs. The replication therefore trained all models for 30 epochs with five seeds and compared each reduced model with the baseline on the six retained features only. Under this matched comparison, omitting any feature still lowered the error on the remaining features, in all five seeds for all seven features (deltas −0.014 to −0.031). The four-unit bottleneck thus reconstructs six features more accurately than the same six features embedded among seven, which is a capacity effect and not evidence that any feature is uninformative.

Table 3. Ablation 2: leave-one-feature-out analysis (original single run versus five-seed matched-feature replication).

Omitted feature	Single-run Δ overall MSE (baseline 0.0772; 25 vs 30 epochs)	5-seed matched-feature Δ MSE (mean ± SD)	Seeds with decrease
l_arm_elev	-0.0162	-0.0309 ± 0.0062	5/5
r_arm_elev	-0.0267	-0.0297 ± 0.0038	5/5
hip_drop_ratio	-0.0226	-0.0284 ± 0.0064	5/5
trunk_tilt	-0.0185	-0.0258 ± 0.0028	5/5
trunk_flexion	-0.0168	-0.0198 ± 0.0041	5/5
l_knee_angle	-0.0160	-0.0169 ± 0.0020	5/5
r_knee_angle	-0.0199	-0.0142 ± 0.0034	5/5

Note. Rows are ordered by the matched-feature replication (largest decrease first). Matched-feature Δ compares each reduced model with the baseline on the six retained features, with all models trained for 30 epochs.

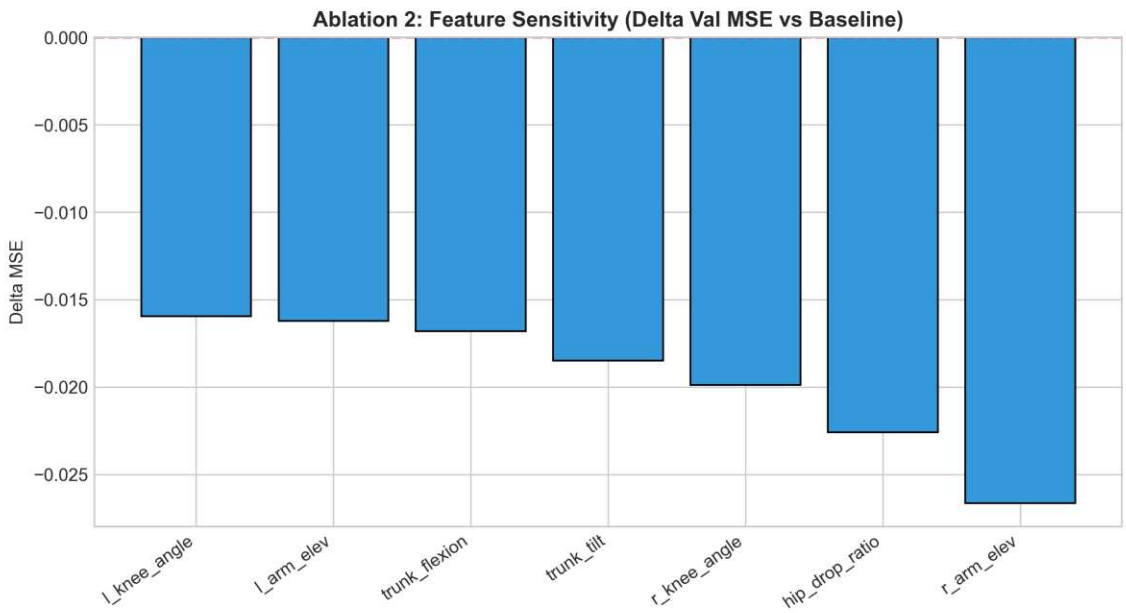


Figure 3. Ablation 2 (original single run): change in overall validation MSE relative to the seven-feature baseline when each feature is omitted (all changes are negative).

The ranking differed from the single run. Omitting left arm elevation (−0.031 ± 0.006), right arm elevation (−0.030 ± 0.004), hip drop ratio (−0.028 ± 0.006), or trunk lateral tilt (−0.026 ± 0.003) relieved the bottleneck most, whereas the knee angles relieved it least (left −0.017 ± 0.002; right −0.014 ± 0.003). This is broadly consistent with the correlation structure of the features: the two knee angles are the most strongly related pair (Spearman $\rho = 0.50$) and can be compressed jointly, whereas arm elevations and trunk tilt are only weakly related to the other features (mean absolute $\rho \leq 0.10$) and each consumes more independent capacity. The right-side-larger asymmetry suggested by the single run was not reproduced, and we draw no inference about handedness.

Figure 4 reports the third ablation, a pairwise correlation matrix of frame-level anomaly predictions from the four detectors on the validation fold. The three classical detectors agreed with one another moderately to strongly (Isolation Forest–One-Class SVM $r = 0.624$; One-Class SVM–PCA reconstruction $r = 0.678$; Isolation Forest–PCA reconstruction $r = 0.591$), whereas the autoencoder correlated more weakly with all

three (r = 0.435 with Isolation Forest, 0.410 with One-Class SVM, and 0.306 with PCA reconstruction). The classical detectors therefore tend to flag a substantially overlapping set of geometrically extreme frames in the scaled feature space, while the non-linear autoencoder identifies a partly distinct set, differing most from the strictly linear PCA baseline. This pattern motivated requiring agreement between the Isolation Forest and the autoencoder rather than among the mutually correlated shallow detectors: a tree-based and a non-linear reconstruction paradigm provide two more nearly independent sources of evidence.

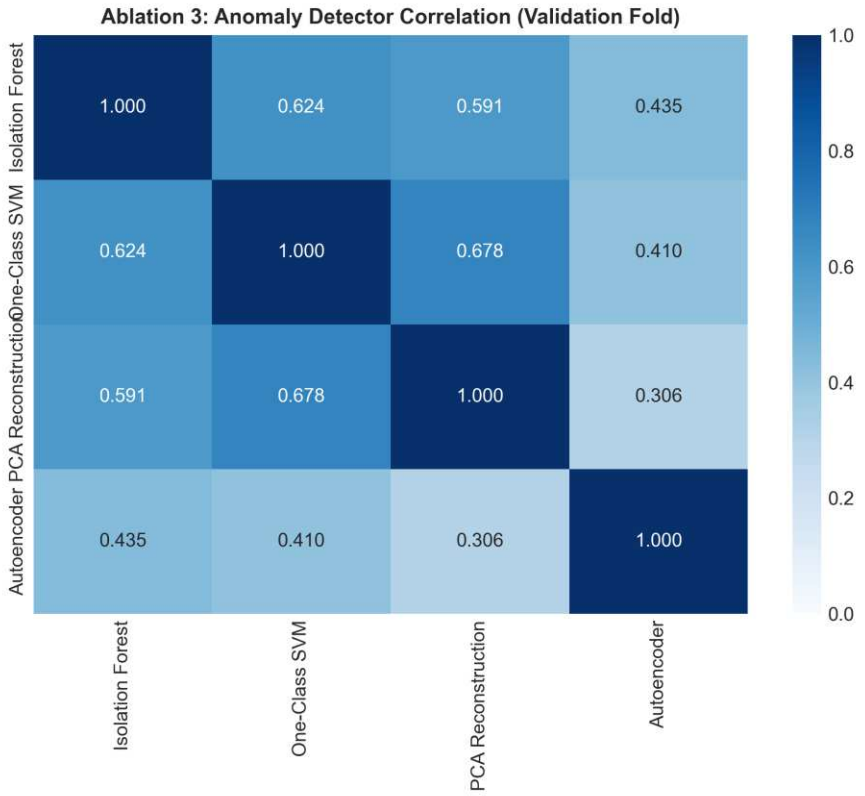


Figure 4. Ablation 3: pairwise correlation of frame-level anomaly predictions across the four detectors on the validation fold.

4.3 Cross-Model Consensus

Figure 5 reports the resulting agreement on the held-out validation split. Of 6,744 frames, the Isolation Forest and the autoencoder jointly classified 5,924 as normal and 256 as anomalous, and disagreed on 564

(215 flagged only by the autoencoder and 349 only by the Isolation Forest), giving Cohen's kappa = 0.431, which lies in the 'moderate' band of the Landis and Koch [42] scale. Given that the two detectors formalise anomaly through unrelated mechanisms, moderate agreement is unsurprising. Applying the same joint criterion to all 36,783 frames, the Isolation Forest flagged 2,107 frames (5.73%), the autoencoder 1,973 (5.36%), and 777 (2.11%) were flagged by both; the kappa on the full corpus was lower (0.345) than on the validation split. These 777 consensus anomalies are the frames carried forward for ergonomic interpretation.

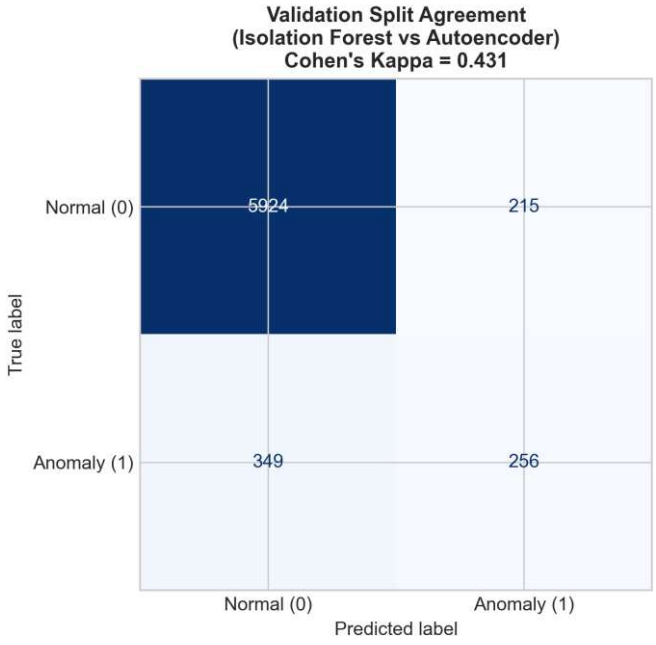


Figure 5. Agreement between Isolation Forest (rows) and autoencoder (columns) anomaly classifications on the held-out validation split (Cohen's $\kappa = 0.431$).

4.4 Latent Posture Structure

Table 4 compares K-Means solutions for k = 2 to 10. The silhouette coefficient was highest at k = 2 (0.231) and similar for k = 3 to 5 (0.212, 0.214, 0.217); the Davies–Bouldin index improved to k = 5 (1.429) and then levelled off, and the Calinski–Harabasz index peaked at k = 4 (7,430.7). Every solution with $k \geq 4$ contained a cluster holding at most 4.9% of frames, whereas at k = 3 the smallest cluster held 21.0%. We

retained k = 3 for interpretability and cluster balance. This is a judgement-based choice and not an index optimum (silhouette alone would have favoured k = 2), so statements about archetypes are conditional on it.

Table 4. K-Means solutions for k = 2 to 10 on the scaled seven-feature corpus (k = 3 retained).

k	Silhouette	Davies–Bouldin (lower better)	Calinski–Harabasz	Smallest cluster (% of frames)
2	0.231	1.997	7,100.5	29.1
3	0.212	1.756	7,266.4	21.0
4	0.214	1.597	7,430.7	4.9
5	0.217	1.429	7,251.7	4.9
6	0.179	1.460	7,076.5	4.7
7	0.183	1.438	6,572.8	2.3
8	0.146	1.579	6,089.7	2.2
9	0.144	1.593	5,726.9	2.2
10	0.144	1.599	5,443.8	2.1

Note. Ten initialisations, seed 42; silhouette on a random 10,000-frame subsample.

Figure 6 shows the corpus-wide posture space, projecting the seven scaled features onto their first two principal components, coloured by K-Means cluster (k = 3) with the 777 consensus anomalies overlaid. Internal validity indices indicated a moderate but genuine cluster structure: a silhouette coefficient of 0.212 (computed on a random 10,000-frame subsample), a Davies–Bouldin index of 1.756, and a Calinski–Harabasz index of 7,266.4 [50, 53, 54]. A silhouette value of this size indicates overlapping clusters, as expected when motor-manual postures form a continuum rather than discrete states [51], and the visible overlap in Figure 6 confirms this. Table 5 gives the physical-unit centroid and size of each cluster. Cluster 0 (20,577 frames; 55.9%) combined modest trunk flexion (22.3°) with near-extended knees (146–148°),

consistent with upright standing, walking, and transitions between cuts. Cluster 1 (8,499 frames; 23.1%) combined moderate trunk flexion (35.2°) with strongly flexed knees (74–83°), consistent with squatting or kneeling while delimbing near the ground. Cluster 2 (7,707 frames; 21.0%) showed the deepest trunk flexion (81.2°) with moderately flexed knees (131°) and the highest arm elevations, consistent with deep stooping during bucking or cross-cutting. These labels are interpretive; no frames were visually annotated to confirm them. The consensus anomalies were strongly concentrated away from Cluster 0: 554 (71.3%) fell in Cluster 2, 210 (27.0%) in Cluster 1, and only 13 (1.7%) in Cluster 0, corresponding to anomaly rates of 7.2%, 2.5%, and 0.06% of each cluster's frames. The reconstruction-based signal is therefore preferentially sensitive to extreme combinations of trunk and limb loading rather than to outliers within the benign standing posture.

Table 5. Posture cluster sizes, physical-unit centroids, centroid OWAS codes (load class 1), consensus-anomaly counts, and interpretive labels.

Cluster	Frames (%)	Trunk flexion (°)	Arm elev. L / R (°)	Knee angle L / R (°)	Hip drop ratio	Consensus anomalies (% of cluster)	Centroid OWAS code (AC)	Interpretive label
0	20,577 (55.9%)	22.3	43.6 / 46.8	148.5 / 146.3	1.44	13 (0.06%)	2131 (AC2)	Upright standing / locomotion
1	8,499 (23.1%)	35.2	49.5 / 43.9	74.3 / 83.4	0.81	210 (2.5%)	2141 (AC2)	Squatting / kneeling (delimbing)

Cluster	Frames (%)	Trunk flexion (°)	Arm elev. L / R (°)	Knee angle L / R (°)	Hip drop ratio	Consensus anomalies (% of cluster)	Centroid OWAS code (AC)	Interpretive label
2	7,707 (21.0%)	81.2	62.6 / 58.1	130.7 / 130.8	1.88	554 (7.2%)	2131 (AC2)	Deep stooping (bucking / cross-cutting)

Note. Centroids are cluster means of the raw (inverse-scaled) features; mean trunk lateral tilt was 0.8°, 0.9°, and 1.7°. Labels are interpretive and were not confirmed by visual annotation. The centroid code under load class 2 is 2132, 2142, and 2132 (AC2, AC3, AC2).

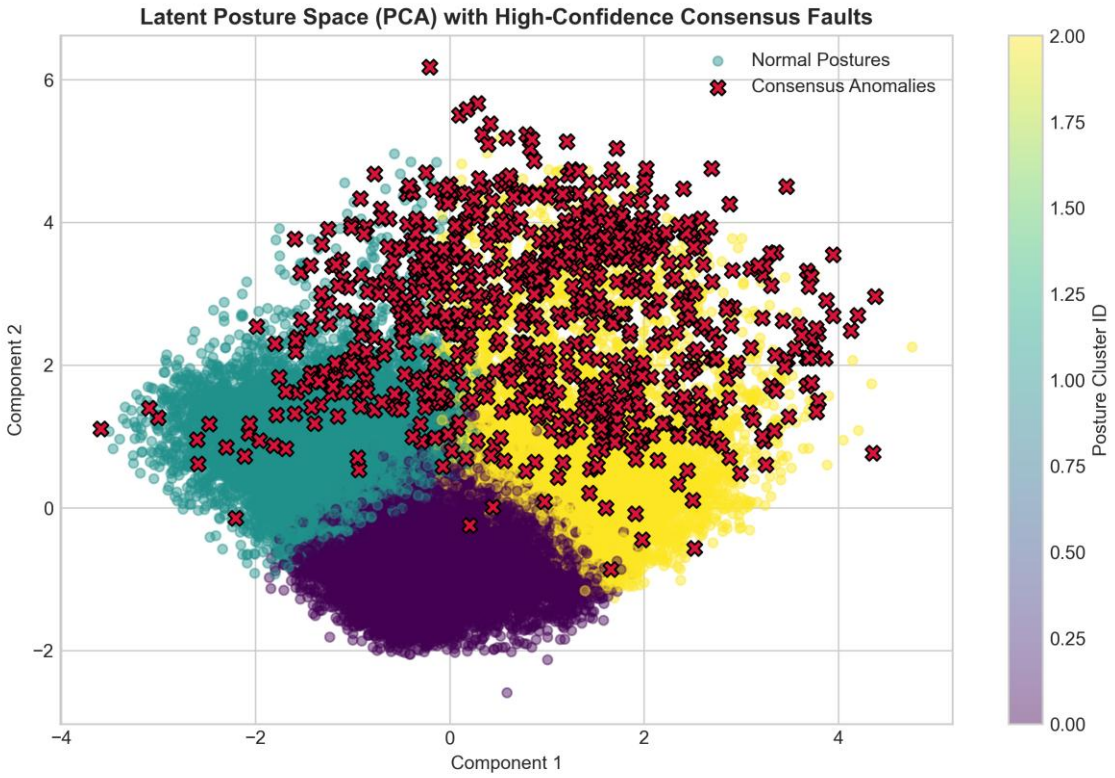


Figure 6. Latent posture space (first two principal components) coloured by K-Means cluster (k = 3), with the 777 consensus anomalies overlaid.

Applying the OWAS rules to the cluster centroids gave codes 2131 for Cluster 0, 2141 for Cluster 1, and 2131 for Cluster 2, all Action Category 2 (Table 5). Centroid-level mapping is therefore coarse: it assigns the same code to the upright cluster (trunk flexion 22.3°, just above the 20° threshold for a bent back) and to the deep-stoop cluster (81.2°), because OWAS has a single bent-back digit. The exposure profiles below accordingly use frame-level coding, which retains the spread within clusters: under frame-level coding 9.3%, 14.5%, and 25.2% of the frames in Clusters 0, 1, and 2 respectively fell in AC3.

4.5 OWAS Classification

Table 6 summarises the distribution of OWAS digits and Action Categories after applying the decision rules of Section 3.8 to every frame. For the back, 12,920 frames (35.1%) were coded straight (digit 1) and 23,863 (64.9%) bent (digit 2). No frame reached the 15° lateral-tilt threshold required for the twisted-back digits (3 or 4), because the maximum trunk lateral tilt in the corpus was 9.74°; since this feature is estimated from a single, mostly fixed camera view, the absence of twisted codes may partly reflect projection limits rather than a true absence of torsion. For the arms, 29,607 frames (80.5%) had both arms below shoulder height (digit 1), 5,559 (15.1%) one arm at or above it (digit 2), and 1,617 (4.4%) both arms elevated (digit 3). For the legs, the distribution was standing (digit 2, 23.3%), uneven or transitional stance (digit 3, 30.9%), squatting (digit 4, 24.4%), one knee bent (digit 5, 15.9%), and kneeling or ground contact (digit 6, 5.6%); no frames were coded as sitting (digit 1) or walking (digit 7). With the load digit fixed at class 1, the OWAS matrix gave 33.3% AC1, 52.8% AC2, and 13.8% AC3 across the corpus. No frame reached AC4. In the matrix used here, AC4 requires either a twisted back (digits 3 or 4) or a load above 20 kg (class 3); because no twisted-back codes occurred and the load digit was at most class 2, AC4 was unreachable, and its absence reflects these structural conditions rather than evidence that extreme postures did not occur.

Table 6. Distribution of OWAS digits and Action Categories across all 36,783 frames, with an upper-bound load-class-2 scenario.

OWAS dimension	Digit / category	Frames (n)	%
Back	1 – Straight	12,920	35.1
Back	2 – Bent	23,863	64.9
Back	3 / 4 – Twisted / bent and twisted	0	0.0
Arms	1 – Both below shoulder	29,607	80.5
Arms	2 – One at/above shoulder	5,559	15.1
Arms	3 – Both at/above shoulder	1,617	4.4
Legs	2 – Standing	8,570	23.3
Legs	3 – Uneven / transitional stance	11,366	30.9
Legs	4 – Squatting	8,970	24.4
Legs	5 – One knee bent	5,834	15.9
Legs	6 – Kneeling / ground contact	2,043	5.6
Load (primary)	1 – Under 10 kg (fixed)	36,783	100.0
Action Category (load class 1, primary)	AC1 – Normal	12,262	33.3
Action Category (load class 1, primary)	AC2 – Slightly harmful	19,438	52.8
Action Category (load class 1, primary)	AC3 – Distinctly harmful	5,083	13.8
Action Category (load class 1, primary)	AC4 – Extremely harmful	0	0.0
Action Category (load class 2, upper bound)	AC1 – Normal	8,956	24.3

OWAS dimension	Digit / category	Frames (n)	%
Action Category (load class 2, upper bound)	AC2 – Slightly harmful	16,987	46.2
Action Category (load class 2, upper bound)	AC3 – Distinctly harmful	10,840	29.5
Action Category (load class 2, upper bound)	AC4 – Extremely harmful	0	0.0

Note. Legs digits 1 (sitting) and 7 (walking) did not occur in the rule-based coding. The load-class-2 scenario assigns class 2 to every frame and is a sensitivity bound, not an estimate.

The load assumption matters for the severity distribution. In the upper-bound scenario in which every frame is assigned load class 2, the corpus-wide distribution shifted to 24.3% AC1, 46.2% AC2, and 29.5% AC3 (Table 6), and the mean per-recording high-risk exposure rose from 11.7% to 23.0% (range 0.0–48.8% versus 0.0–24.2%). Still no frame reached AC4. The primary results (load class 1) should therefore be read as the lower bound of the plausible severity range, and the profile in the remainder of this section uses load class 1.

4.6 Ergonomic Exposure Profiles

Figure 7 and Table 7 report the per-recording exposure profiles, ranked by the percentage of frames in AC3 + AC4. Exposure was heterogeneous across the 35 recordings (mean 11.7%, range 0.0–24.2%): the highest-exposure recording (IMG_1532.MOV, 455 frames) spent 24.2% of frames in AC3, whereas two recordings (GOPR1074.MP4 and IMG_1830.MOV) had no AC3 frames. The lowest-exposure recordings were also among the shortest (GOPR1074.MP4 contributed only four analysed frames), and recording length varied from 4 to 4,161 frames, so per-recording percentages from short clips should be interpreted with caution and reliable per-worker estimates need a minimum sampled duration. Figure 7 shows only the 15 highest-exposure recordings; Table 7 adds the lowest-exposure recordings and the recording with the highest

anomaly density. The consensus-anomaly density did not track the AC exposure ranking. IMG_1829.MOV ranked 25th of 35 on AC3 + AC4 exposure (8.8%) yet had by far the highest anomaly density (22.2%, 108 of 487 frames), roughly four times that of the next-highest recording (5.4%). This indicates that the reconstruction-based signal responds to frames that the static OWAS code does not single out; we did not review the footage frame by frame, so we cannot say whether these frames correspond to rapid corrections, atypical manoeuvres, tracking artefacts, or other causes. The divergence nonetheless shows that the OWAS exposure layer and the anomaly layer of the framework (Figure 1) carry different information: the former expresses cumulative postural load in a standardised unit, the latter flags frames that depart from the learned normative manifold.

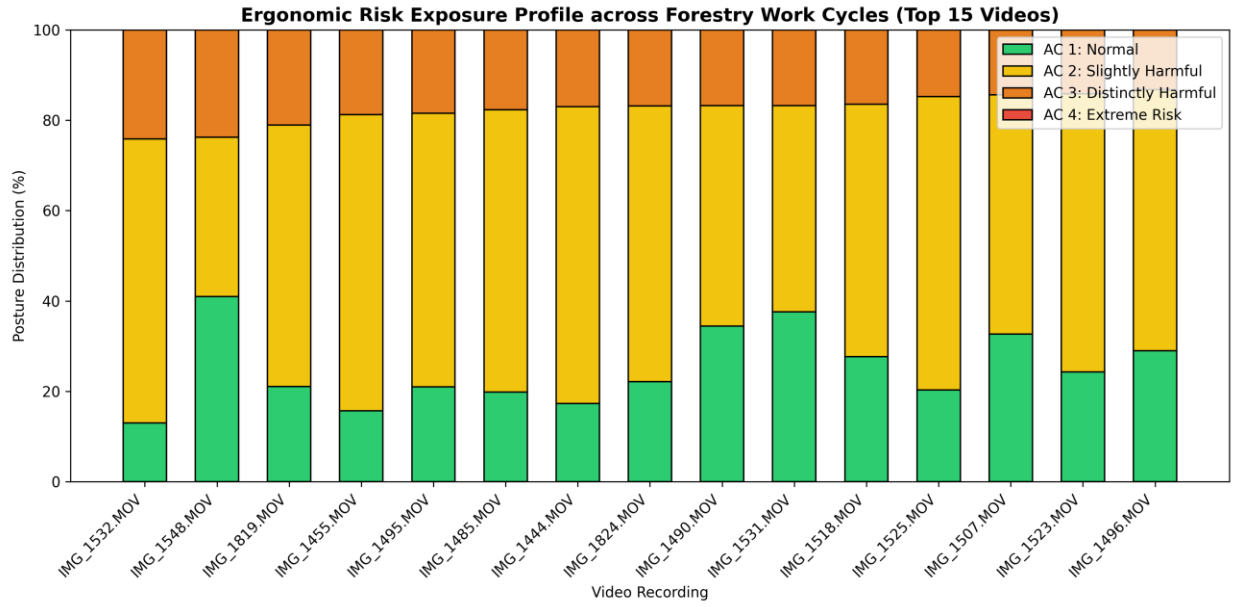


Figure 7. Ergonomic risk exposure profile (percentage of frames per OWAS Action Category, load class 1) for the 15 highest-exposure recordings.

Table 7. Per-recording exposure (load class 1): highest-exposure recordings (ranks 1–10), the highest-anomaly-density recording (rank 25), and the lowest-exposure recordings (ranks 31–35).

Rank	Recording	Frames	AC1 %	AC2 %	AC3 %	AC3+AC4 %	Consensus anomalies	Anomaly density %
1	IMG_1532.MOV	455	13.0	62.9	24.2	24.2	9	2.0
2	IMG_1548.MOV	122	41.0	35.2	23.8	23.8	0	0.0
3	IMG_1819.MOV	57	21.1	57.9	21.1	21.1	3	5.3
4	IMG_1455.MOV	709	15.7	65.6	18.8	18.8	34	4.8
5	IMG_1495.MOV	276	21.0	60.5	18.5	18.5	0	0.0
6	IMG_1485.MOV	2944	19.9	62.4	17.7	17.7	74	2.5
7	IMG_1444.MOV	1519	17.3	65.7	17.0	17.0	41	2.7
8	IMG_1824.MOV	208	22.1	61.1	16.8	16.8	1	0.5
9	IMG_1490.MOV	4161	34.4	48.8	16.8	16.8	57	1.4
10	IMG_1531.MOV	1571	37.6	45.7	16.7	16.7	33	2.1
25	IMG_1829.MOV	487	28.8	62.4	8.8	8.8	108	22.2
31	IMG_1827.MOV	428	73.1	23.8	3.0	3.0	2	0.5
32	IMG_1822.MOV	67	88.1	10.4	1.5	1.5	0	0.0
33	IMG_1508.MOV	178	92.1	7.3	0.6	0.6	0	0.0
34	GOPR1074.MP4	4	100.0	0.0	0.0	0.0	0	0.0
35	IMG_1830.MOV	117	100.0	0.0	0.0	0.0	0	0.0

Note. No frame reached AC4, so AC3 + AC4 equals AC3. Recordings are ranked by AC3 + AC4 exposure across all 35 recordings.

4.7 Summary of Findings

Taken together, the results are consistent with the four propositions advanced in Section 2.3, subject to the caveats stated above. Consistent with Proposition 1, the four-unit autoencoder converged on held-out videos (validation MSE 0.0886 in the original run and 0.0819 ± 0.0022 across five seeds), and the four-unit bottleneck gave the best separation between Isolation Forest-flagged and unflagged frames among the sizes tested. Consistent with Proposition 2, the Isolation Forest and the autoencoder agreed beyond chance but only moderately (kappa = 0.431 on the validation split, 0.345 on the full corpus), and the autoencoder was the detector least correlated with the others (Figure 4), so the 777-frame consensus set is a markedly more conservative selection than either detector's own flags (2,107 and 1,973 frames). Consistent with Proposition 3, three overlapping but interpretable archetypes emerged (silhouette = 0.212), and consensus anomalies were concentrated in the two flexed-posture clusters rather than in the upright cluster (Table 5), although k = 3 was a judgement-based choice among near-equivalent solutions (Table 4). Consistent with Proposition 4, anomaly density and AC3 + AC4 exposure ranked recordings differently, with IMG_1829.MOV as the clearest divergence (Table 7).

Three results were structural rather than behavioural and should not be over-interpreted. The absence of AC4 follows from the OWAS matrix (it requires a twisted back or a load above 20 kg) together with a maximum lateral tilt of 9.74° against a 15° threshold in single-view footage; the severity distribution depends materially on the load assumption (13.8% versus 29.5% AC3 under load class 1 versus 2); and the very low exposure of the shortest recordings reflects small samples. None of these results speaks to the true ergonomic hazard of the work, which can only be established against external references.

5. Discussion

5.1. Interpretation of the Findings

This study asked whether a fully unsupervised, video-based pipeline could quantify OWAS ergonomic risk exposure in authentic motor-manual forestry footage without manually labelled ground truth. The results

address the three research questions with important qualifications. For RQ1, a compact four-unit autoencoder converged on held-out videos under a strict group-based split (Table 2, Figure 2), and the matched, multi-seed leave-one-feature-out analysis (Table 3, Figure 3) showed that every feature adds to the demand on the four-unit bottleneck, most for the weakly correlated arm and trunk-tilt variables, with a feature ranking that differed from the original single run. For RQ2, tree-based isolation and non-linear reconstruction agreed only moderately (kappa = 0.431; Figure 5), and the correlation ablation (Figure 4) showed that the autoencoder is least similar to the classical detectors; no single unsupervised detector should therefore be treated as ground truth, in line with the reproducibility concerns that motivate multi-model, leakage-audited validation [29]. For RQ3, the clusters and consensus anomalies mapped onto OWAS in a way that is plausible for the task, with a three-archetype repertoire of upright transitional standing, squat-based delimbing, and stooped bucking (Table 5, Figure 6) and an exposure profile of 33.3% AC1, 52.8% AC2, and 13.8% AC3 (Table 6). We did not compare this profile with manually coded OWAS data from the same recordings, so the claim is one of plausibility and internal consistency, not of accuracy.

These findings extend a computer-vision ergonomics literature that has concentrated on supervised or semi-supervised posture classification, typically validated on construction sites, assembly lines, farm tasks, and laboratory simulations [15–24]. The most directly comparable forestry work trained deep-learning models on thousands of manually annotated cross-cutting images to reproduce OWAS classes [4]. The present pipeline instead discovers posture structure from the unlabelled geometry of the data, at the cost of coarser, internally validated resolution (three archetypes rather than fine-grained OWAS classes) and with the benefit of requiring no annotation, which matters more as the number of workers, sites, and hours monitored grows. It is therefore best viewed as a low-annotation-cost screening layer rather than a competitor to supervised OWAS classifiers: unsupervised exposure profiling could flag the highest-exposure segments and workers (Figure 7, Table 7) for supervised review or on-site consultation, echoing the augmenting, human-centric vision articulated for Industry 5.0 [30].

Our kappa values should also be read against the reliability of OWAS itself. In a forestry study of human and machine OWAS ratings, agreement between individual human raters and a deep-learning model ranged from poor to fair (Cohen's kappa −0.03 to 0.34), rising to fair-to-moderate when all raters were pooled with the model [6]. Our detector-to-detector kappa of 0.431 is of a similar order, although it measures agreement between two algorithms on anomaly status rather than between raters on OWAS codes, so the comparison is only indicative. It does suggest that moderate agreement is not unusual in this domain, where posture varies along a continuum and any discretisation is partly arbitrary.

The divergence between OWAS exposure and anomaly density is, in our view, a practically important finding. A categorical code is by design insensitive to transitions between postures and to brief excursions within an otherwise moderate work cycle, whereas a reconstruction-based signal is tuned to frames that depart from the learned normative manifold. A recording such as IMG_1829.MOV, unremarkable on AC3 + AC4 exposure but exceptional on anomaly density, would pass a conventional OWAS audit without comment. Whether its anomalous frames mark genuinely hazardous events, unusual but benign movements, or pose-tracking artefacts is an empirical question that the present data cannot settle; this is precisely why the framework's highlighted verification video exists, so that a human reviewer can inspect flagged frames. We therefore suggest that anomaly detection and standardised observational coding be reported side by side, as proposed here, rather than either substituting for the other.

5.2. Practical Implications

For forestry employers, safety managers, and occupational health services, the main practical value of the pipeline lies in prioritisation. A manual OWAS audit of every recording is rarely affordable, but a label-free first pass can rank recordings by high-risk exposure and by anomaly density, directing scarce expert attention to the recordings that matter most. In this corpus, for example, the ten highest-exposure recordings spent between 16.7% and 24.2% of analysed frames in AC3, compared with a corpus mean of 11.7% per

recording (Table 7), and a reviewer who inspected only those recordings, and the high-anomaly recording IMG_1829.MOV, would cover the most informative material at a fraction of the cost of reviewing all 35.

Because the OWAS output is expressed as the share of frames in each Action Category, it translates directly into the language already used in ergonomic reports and in risk-management documentation. The measure can be tracked over time for a crew, compared before and after an intervention such as a change in cutting technique, work-rest scheduling, or tool configuration, and used to show whether a particular task element (here, the flexed-posture clusters that contain most of the anomalies) is where exposure concentrates. The automatically rendered verification video supports supervisor feedback and training, because it lets a worker see the flagged moment in context without needing to interpret any model output. These uses are consistent with the human-centric orientation of Industry 5.0, in which digital tools are intended to augment rather than replace the judgement of workers and safety professionals [30].

A responsible deployment would also set minimum-data rules. Our results show that recordings of a few dozen frames can produce extreme percentages (0% or 100% in the AC categories) that carry little information, so exposure summaries should be reported with the number of analysed frames and, for operational use, only for recordings or aggregated worker-days above a stated minimum duration. Aggregating across recordings by worker, task, and shift, rather than reporting per clip, would make the estimates more stable and more relevant to the cumulative-exposure logic on which ergonomic risk is based.

5.3. Theoretical and Methodological Implications

Theoretically, the study supports the view that observational ergonomics and unsupervised representation learning are complementary rather than competing traditions. OWAS contributes an interpretable, intervention-oriented endpoint but a coarse discretisation; reconstruction- and isolation-based detectors contribute sensitivity to deviation without any claim about what the deviation means. The divergence between AC exposure and anomaly density in our data is a concrete illustration of that complementarity,

and it suggests that future ergonomic surveillance frameworks should report both a cumulative-category measure and a deviation measure rather than collapsing them into one score.

Methodologically, the study offers a template for reporting leakage control in video-frame studies: grouping by source recording, fitting preprocessing and thresholds on training data only, quarantining non-kinematic metadata, and reporting the audit outcome. These practices address the concern, raised across many scientific fields, that unaudited leakage inflates reported performance and undermines reproducibility [29]. The study also illustrates a useful discipline in unsupervised work: because no ground truth exists, claims were restricted to internal validity indices, cross-model agreement, and plausibility against the task, and each of these was reported together with its limits. The moderate silhouette and kappa values we report are not failures of the method; they describe how continuous the underlying postural space is, and they should be expected in any similar attempt to discretise it.

5.4. Ethical and Privacy Considerations

Continuous video monitoring of workers raises legitimate concerns about privacy, consent, and the potential misuse of exposure data for performance management rather than protection. The footage analysed here was used solely to extract anonymous kinematic features, and the pipeline outputs only posture-derived quantities; no identifiable image data leave the analysis. In operational use, we recommend that monitoring be introduced with the informed involvement of workers and their representatives, that exposure data be used to improve tasks, tools, and schedules rather than to rank individuals disciplinarily, that retention of raw video be minimised, and that flagged frames be reviewed by a trained person before any action is taken, since the anomaly flag is a statistical signal and not a finding of unsafe behaviour. These recommendations follow the logic of human-centric AI in the Industry 5.0 literature, which treats worker well-being and agency as design objectives rather than constraints [30].

5.5. Limitations and Future Research

Several limitations qualify these conclusions. First, no ground truth exists: the pipeline was not validated against expert-coded OWAS labels, marker-based motion capture, or visual review of flagged frames, so absolute accuracy of the joint angles, clusters, OWAS codes, and anomaly flags is unknown, and monocular, single-view pose estimation in an outdoor forest is vulnerable to occlusion and projection error even with a 0.5 visibility filter. Second, the load digit was fixed at class 1 for every frame, so the pipeline cannot distinguish unloaded repositioning from active cutting; the upper-bound scenario shows that this assumption alone moves the AC3 share between 13.8% and 29.5%, so the absolute exposure percentages are conditional on it, whereas the ranking of recordings is less affected. Third, the absence of twisted-back codes may reflect camera-view limits rather than true torsion, and together with the load assumption it makes AC4 unreachable; multi-view or depth-aware pose estimation could sharpen this dimension. Fourth, the ablations in Figures 2 and 3 come from a single PyTorch run, supplemented by a five-seed replication in an independent NumPy implementation; the replication reproduced the capacity trend but not the single-run batch-normalisation result or feature ranking, and its separation test uses Isolation Forest flags rather than true anomalies as the reference. Fifth, the moderate silhouette (0.212) and kappa (0.431) indicate genuine but imperfect structure, the choice of $k = 3$ was judgement-based, and K-Means was fitted on all frames, so cluster labels for validation frames were not predicted out of sample; frame-independent modelling also ignores the strong short-range autocorrelation of human movement, and temporal (sequence-aware) models, density-based clustering [59], or stronger regularisation such as dropout [60] might improve separation. Finally, the sample of 35 recordings, with highly unequal lengths and an unreported number of operators, should be extended across more operators, seasons, terrain types, and sub-tasks before the specific exposure percentages are generalised.

The limitations above define a concrete programme of future work in roughly increasing order of cost. The cheapest and most informative step is validation against external references: expert-coded OWAS labels on a stratified sample of frames, including consensus anomalies and an equal number of non-anomalies, would

quantify how well the unsupervised archetypes and anomaly flags correspond to human judgement, and a review of the highest-density recordings would show what the anomalous frames contain. A second step is to replace the fixed load digit with a dynamic one driven by a tool-engagement classifier, which would remove the present dependence on a bounding scenario. A third is to repeat the ablations with the original PyTorch code over multiple seeds, with confidence intervals and a normal-versus-anomalous separation test against an independent reference, and to out-of-sample-validate the clustering by fitting K-Means on the training partition only. A fourth is methodological extension: sequence-aware models that exploit the temporal structure of movement, multi-view or depth-aware pose estimation to recover trunk rotation, and comparison against alternative coding schemes such as REBA and RULA, which weight the upper limbs more heavily than OWAS [7, 8]. Finally, deployment studies that follow crews over entire shifts and seasons, with the privacy safeguards described above, would test whether the exposure measures change in response to interventions, which is the ultimate test of the usefulness of any ergonomic surveillance system.

6. Conclusions

This study developed a fully unsupervised, leakage-audited, video-based pipeline that converts raw motor-manual forestry footage into quantitative OWAS exposure profiles without manually labelled posture data. Combining classical detectors with a compact deep autoencoder, and requiring agreement between the Isolation Forest and the autoencoder before a frame was treated as a consensus anomaly, the pipeline identified a coherent three-archetype posture repertoire and an Action Category distribution (33.3% AC1, 52.8% AC2, 13.8% AC3) that is plausible for the task, while also surfacing frames that static posture coding does not single out. The principal contribution is methodological: the work shows that corpus-scale, label-free ergonomic screening of a high-risk field occupation is feasible, with explicit leakage control, ablation-based diagnostics, and a translation into a standardised, practitioner-interpretable coding scheme. The framework is in principle transferable to other occupations characterised by unscripted, video-recordable manual labour. For practice, per-recording exposure rankings and the highlighted verification video can help prioritise which workers, tasks, or periods merit closer ergonomic attention, as a complement to rather

than a replacement for expert observational audits. For research, the principal next steps are validation against expert-coded OWAS labels and motion capture, dynamic load inference, multi-view or depth-aware pose estimation, replicated and seed-controlled ablations, and temporally aware unsupervised architectures.

Table 8. Table of abbreviations and acronyms

Table 8 defines all the abbreviations and acronyms used in this study for readers to easily understand the content of this study.

Abbreviation/Acronym	**Full Meaning**
Institutional & Sectoral	
BLS	Bureau of Labor Statistics (United States)
NIOSH	National Institute for Occupational Safety and Health
Ergonomics & Occupational Health	
AC	Action Category (OWAS severity index; AC1–AC4)
OWAS	Ovako Working Posture Analysing System
REBA	Rapid Entire Body Assessment
RULA	Rapid Upper Limb Assessment
WMSD	Work-related musculoskeletal disorder
Computer Vision, Video & Artificial Intelligence	
AI	Artificial intelligence
AVI	Audio Video Interleave (video file format)
MOV	QuickTime movie (video file format)
MP4	MPEG-4 Part 14 (video file format)
RMPE	Regional Multi-Person Pose Estimation
Machine Learning, Statistics & Software	

AdamW	Adam optimiser with decoupled weight decay
AUC	Area under the receiver operating characteristic curve
BN	Batch normalisation
LeakyReLU	Leaky rectified linear unit (activation function)
MSE	Mean squared error
NumPy	Numerical Python (scientific computing library)
PCA	Principal Component Analysis
ROC	Receiver operating characteristic
RQ	Research question
SVM	Support vector machine
κ	Cohen's kappa coefficient of agreement

Declarations

Ethics approval and consent to participate: Not applicable

Consent for publication: Not applicable

Availability of data and materials: Data and materials that support the research findings and code snippets used to train the models, analyze data and create the visualizations are available at: https://doi.org/10.6084/m9.figshare.35082742.

Competing interests: The author has no relevant financial or non-financial interests to disclose.

Funding: This work received no funding.

Authors' contributions:

G.O.F.: Conceptualization, Methodology, Formal analysis, Data curation, Validation, Investigation, Visualization, Writing – original draft, Writing – editing & revision, Resources, Supervision, Project administration.

F.A-K: Conceptualization, Methodology, Formal analysis, Data curation, Validation, Investigation, Visualization, Writing – original draft, Writing – editing & revision, Resource, Project administration.

Declaration of Generative AI and AI-assisted Technologies in the Manuscript Preparation Process: During the preparation of this work, the authors used AI-assisted technology, namely Cursor version 2.4.37, to write and edit the Python code used to train the models, analyze data and create visualizations in this study. After using this tool/service, the author reviewed and edited the content as needed and take full responsibility for the content of the published article.

Acknowledgements: The authors thank the Department of Forest Engineering, Forest Management Planning, and Terrestrial Measurements, Faculty of Silviculture and Forest Engineering, Transilvania University of Brasov, and the Forestry Commission of Ghana, Forest Services Division, Accra for providing some of the resources needed for this study.

References

1. National Institute for Occupational Safety and Health. Perspectives on forest operations safety. NIOSH Science Blog, Centers for Disease Control and Prevention; 2024.
2. Bureau of Labor Statistics. National census of fatal occupational injuries in 2023. Washington: U.S. Department of Labor; 2024.

3. Corella Justavino F, Jimenez Ramirez R, Meza Perez N, Borz SA. The use of OWAS in forest operations postural assessment: advantages and limitations. Bull Transilv Univ Brasov Ser II For Wood Ind Agric Food Eng. 2015;8:7-16.

4. Forkuo GO, Borz SA, Kaakkurivaara T, Kaakkurivaara N. Postural classification by image embedding and transfer learning: an example of using the OWAS method in motor-manual work to automate the process and save resources. Forests. 2025;16:492. doi:10.3390/f16030492.

5. Karhu O, Kansi P, Kuorinka I. Correcting working postures in industry: a practical method for analysis. Appl Ergon. 1977;8:199-201. doi:10.1016/0003-6870(77)90164-8.

6. Forkuo GO, Marcu MV, Kaakkurivaara N, Kaakkurivaara T, Borz SA. Human and machine reliability in postural assessment of forest operations by OWAS method: level of agreement and time resources. Forests. 2025;16:759. doi:10.3390/f16050759.

7. Hignett S, McAtamney L. Rapid entire body assessment (REBA). Appl Ergon. 2000;31:201-5. doi:10.1016/S0003-6870(99)00039-3.

8. McAtamney L, Corlett EN. RULA: a survey method for the investigation of work-related upper limb disorders. Appl Ergon. 1993;24:91-9. doi:10.1016/0003-6870(93)90080-S.

9. Kong Y-K, Lee S-Y, Lee K-S, Kim D-M. Comparisons of ergonomic evaluation tools (ALLA, RULA, REBA and OWAS) for farm work. Int J Occup Saf Ergon. 2018;24:218-23. doi:10.1080/10803548.2017.1306960.

10. Micheletti Cremasco M, Giustetto A, Caffaro F, Colantoni A, Cavallo E, Grigolato S. Risk assessment for musculoskeletal disorders in forestry: a comparison between RULA and REBA in the manual feeding of a wood-chipper. Int J Environ Res Public Health. 2019;16:793. doi:10.3390/ijerph16050793.

11. Diego-Mas JA, Alcaide-Marzal J. Using Kinect sensor in observational methods for assessing postures at work. Appl Ergon. 2014;45:976-85. doi:10.1016/j.apergo.2013.12.001.

12. Cao Z, Simon T, Wei S-E, Sheikh Y. Realtime multi-person 2D pose estimation using part affinity fields. In: Proceedings of the IEEE conference on computer vision and pattern recognition. Piscataway: IEEE; 2017. p. 7291-9. doi:10.1109/CVPR.2017.143.

13. Fang H-S, Xie S, Tai Y-W, Lu C. RMPE: regional multi-person pose estimation. In: Proceedings of the IEEE international conference on computer vision. Piscataway: IEEE; 2017. p. 2334-43. doi:10.1109/ICCV.2017.256.

14. Lugaresi C, Tang J, Nash H, McClanahan C, Uboweja E, Hays M, et al. MediaPipe: a framework for building perception pipelines. arXiv:1906.08172. 2019; doi:10.48550/arXiv.1906.08172.

15. Yan X, Li H, Wang C, Seo J, Zhang H, Wang H. Development of ergonomic posture recognition technique based on 2D ordinary camera for construction hazard prevention through view-invariant features in 2D skeleton motion. Adv Eng Inform. 2017;34:152-63. doi:10.1016/j.aei.2017.11.001.

16. Zhao J, Obonyo E. Applying incremental deep neural networks-based posture recognition model for ergonomics risk assessment in construction. Adv Eng Inform. 2021;50:101374. doi:10.1016/j.aei.2021.101374.

17. Chatzis T, Konstantinidis D, Dimitropoulos K. Automatic ergonomic risk assessment using a variational deep network architecture. Sensors. 2022;22:6051. doi:10.3390/s22166051.

18. Lu O, Hochgraf C. Deep-learning based automatic ergonomic assessment using webcam data. J Stud Res. 2023;12: doi:10.47611/jsrhs.v12i4.5240.

19. Aghamohammadi A, Beheshti Shirazi SA, Banihashem SY, Shishechi S, Ranjbarzadeh R, Jafarzadeh Ghoushchi S, et al. A deep learning model for ergonomics risk assessment and sports and health monitoring in self-occluded images. Signal Image Video Process. 2024;18:1161-73. doi:10.1007/s11760-023-02830-6.

20. Malaisé A, Maurice P, Colas F, Ivaldi S. Activity recognition for ergonomics assessment of industrial tasks with automatic feature selection. IEEE Robot Autom Lett. 2019;4:1132-9. doi:10.1109/LRA.2019.2894389.

21. Akanmu AA, Olayiwola J, Ogunseiju O, McFeeters D. Cyber-physical postural training system for construction workers. Autom Constr. 2020;117:103272. doi:10.1016/j.autcon.2020.103272.

22. Antwi-Afari MF, Qarout Y, Herzallah R, Anwer S, Umer W, Zhang Y, et al. Deep learning-based networks for automated recognition and classification of awkward working postures in construction using wearable insole sensor data. Autom Constr. 2022;136:104181. doi:10.1016/j.autcon.2022.104181.

23. Zhao J, Obonyo E, Bilén SG. Wearable inertial measurement unit sensing system for musculoskeletal disorders prevention in construction. Sensors. 2021;21:1324. doi:10.3390/s21041324.

24. Li L, Martin T, Xu X. A novel vision-based real-time method for evaluating postural risk factors associated with musculoskeletal disorders. Appl Ergon. 2020;87:103138. doi:10.1016/j.apergo.2020.103138.

25. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science. 2006;313:504-7. doi:10.1126/science.1127647.

26. Zhou C, Paffenroth RC. Anomaly detection with robust deep autoencoders. In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining. New York: ACM; 2017. p. 665-74. doi:10.1145/3097983.3098052.

27. Chandola V, Banerjee A, Kumar V. Anomaly detection: a survey. ACM Comput Surv. 2009;41:15. doi:10.1145/1541880.1541882.

28. Pang G, Shen C, Cao L, van den Hengel A. Deep learning for anomaly detection: a review. ACM Comput Surv. 2021;54:38. doi:10.1145/3439950.

29. Kapoor S, Narayanan A. Leakage and the reproducibility crisis in machine-learning-based science. Patterns. 2023;4:100804. doi:10.1016/j.patter.2023.100804.

30. Mentzas G, Hribernik K, Stahre J, Romero D, Soldatos J. Editorial: human-centered artificial intelligence in Industry 5.0. Front Artif Intell. 2024;7:1429186. doi:10.3389/frai.2024.1429186.

31. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature. 1986;323:533-6. doi:10.1038/323533a0.

32. Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge: MIT Press; 2016.

33. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436-44. doi:10.1038/nature14539.

34. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Proceedings of the 32nd international conference on machine learning. PMLR; 2015. p. 448-56.

35. Maas AL, Hannun AY, Ng AY. Rectifier nonlinearities improve neural network acoustic models. In: Proceedings of the ICML workshop on deep learning for audio, speech and language processing. 2013.

36. Loshchilov I, Hutter F. Decoupled weight decay regularization. In: International conference on learning representations. ICLR; 2019.

37. Kingma DP, Ba J. Adam: a method for stochastic optimization. In: Proceedings of the 3rd international conference on learning representations. ICLR; 2015. doi:10.48550/arXiv.1412.6980.

38. Liu FT, Ting KM, Zhou ZH. Isolation forest. In: 2008 eighth IEEE international conference on data mining. Piscataway: IEEE; 2008. p. 413-22. doi:10.1109/ICDM.2008.17.

39. Breiman L. Random forests. Mach Learn. 2001;45:5-32. doi:10.1023/A:1010933404324.

40. Schölkopf B, Platt JC, Shawe-Taylor J, Smola AJ, Williamson RC. Estimating the support of a high-dimensional distribution. Neural Comput. 2001;13:1443-71. doi:10.1162/089976601750264965.

41. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20:37-46. doi:10.1177/001316446002000104.

42. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159-74. doi:10.2307/2529310.

43. Lloyd SP. Least squares quantization in PCM. IEEE Trans Inf Theory. 1982;28:129-37. doi:10.1109/TIT.1982.1056489.

44. MacQueen J. Some methods for classification and analysis of multivariate observations. In: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, vol. 1. Berkeley: University of California Press; 1967. p. 281-97.

45. Arthur D, Vassilvitskii S. k-means++: the advantages of careful seeding. In: Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms. Philadelphia: Society for Industrial and Applied Mathematics; 2007. p. 1027-35.

46. Hotelling H. Analysis of a complex of statistical variables into principal components. J Educ Psychol. 1933;24:417-41. doi:10.1037/h0071325.

47. Jolliffe IT. Principal component analysis. 2nd ed. New York: Springer; 2002. doi:10.1007/b98835.

48. Pearson K. On lines and planes of closest fit to systems of points in space. Lond Edinb Dubl Phil Mag J Sci. 1901;2:559-72. doi:10.1080/14786440109462720.

49. Ding C, He X. K-means clustering via principal component analysis. In: Proceedings of the twenty-first international conference on machine learning. New York: ACM; 2004. p. 29. doi:10.1145/1015330.1015408.

50. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math. 1987;20:53-65. doi:10.1016/0377-0427(87)90125-7.

51. Shahapure KR, Nicholas C. Cluster quality analysis using silhouette score. In: 2020 IEEE 7th international conference on data science and advanced analytics. Piscataway: IEEE; 2020. p. 747-8. doi:10.1109/DSAA49011.2020.00096.

52. Gwak G-T, Hwang U-J, Kim J-H. Clustering of shoulder movement patterns using K-means algorithm based on the shoulder range of motion. J Bodyw Mov Ther. 2025;41:164-70. doi:10.1016/j.jbmt.2024.11.034.

53. Davies DL, Bouldin DW. A cluster separation measure. IEEE Trans Pattern Anal Mach Intell. 1979;PAMI-1:224-7. doi:10.1109/TPAMI.1979.4766909.

54. Caliński T, Harabasz J. A dendrite method for cluster analysis. Commun Stat Theory Methods. 1974;3:1-27. doi:10.1080/03610927408827101.

55. Hubert L, Arabie P. Comparing partitions. J Classif. 1985;2:193-218. doi:10.1007/BF01908075.

56. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York: Springer; 2009. doi:10.1007/978-0-387-84858-7.

57. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825-30.

58. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: an imperative style, high-performance deep learning library. Adv Neural Inf Process Syst. 2019;32:8026-37.

59. Ester M, Kriegel H-P, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the second international conference on knowledge discovery and data mining. Menlo Park: AAAI Press; 1996. p. 226-31.

60. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res. 2014;15:1929-58.

997

998